# Computer Science Department

# TECHNICAL REPORT

Inferring Ignorance from the Locality
of Visual Perception

*Ernest Davis*

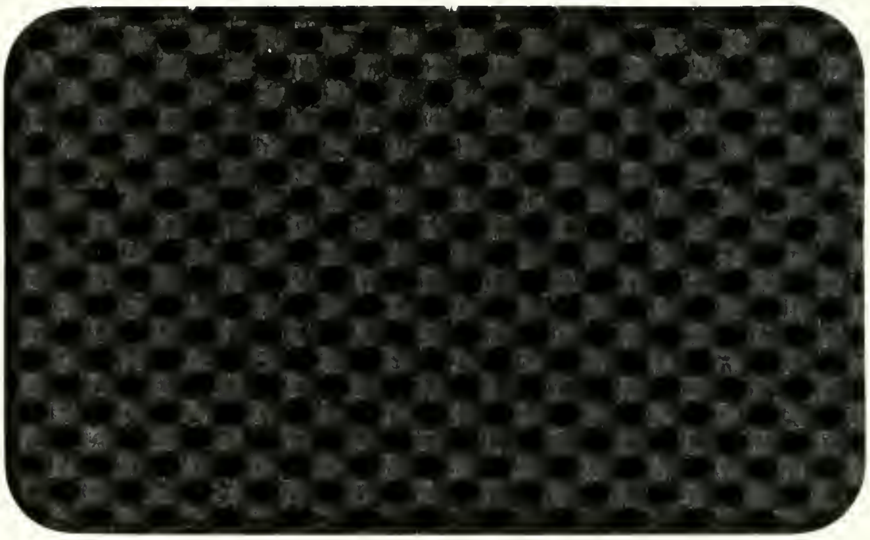
---

Technical Report 328

November 1987


NEW YORK UNIVERSITY

Department of Computer Science
Courant Institute of Mathematical Sciences
251 MERCER STREET, NEW YORK, N.Y. 10012



# Inferring Ignorance from the Locality of Visual Perception

*Ernest Davis*

---

Technical Report 328

November 1987

# Inferring Ignorance from the Locality of Visual Perception

*Ernest Davis*

*ABSTRACT*

It is often possible to infer that a certain agent does not know about a certain occurrence, because he has had no way to find out about it. We present a formal model of knowledge acquisition through visual perception that supports inferences of this kind. The chief innovation of this model is that it distinguishes between physically possible states of the world and epistemically possible worlds. Such a model allows us to state the following axiom: If a course of events is physically possible, and an agent's perceptions are consistent with that course of events, then the agent cannot know that these events did not, in fact, occur. We show how this axiom supports the desired inferences in a simple models of physics and vision. Our model is rich enough to support as well a variety of positive inferences, in which an agent gains knowledge of various kinds through his perception.

November 10, 1987

# Inferring Ignorance from the Locality of Visual Perception

*Ernest Davis*

## 1. Introduction

Intelligent creatures learn much of what they know through direct perception. Therefore, reasoning about the acquisition of knowledge over time often requires a high-level understanding of the power and limits of perception. This paper presents a logical theory that supports high-level reasoning about knowledge and perception. We construct a formal language in which perception can be described. Using this language, we state some fundamental axioms, and we show that these are sufficient to justify some elementary but interesting inferences about perceptions.

One particular type of reasoning that may be supported by a theory of perception is the inference that an agent must be ignorant of a particular fact because he has had no way to find out whether it is true. Such means of inferring ignorance may be important, either to infer that a secret may be safely kept from another party, or to guide an agent in planning a method to find out a given fact. Previous theories of knowledge have not supported any strong methods of inferring ignorance, other than a backward use of consequential closure: If $A$ does not know $Q$ but $A$ does know that $P$ implies $Q$, then $A$ must not know $P$. A theory of knowledge can thus be made a much more powerful tool when it is associated with a theory of perception.

We will focus on the following problem:

I. Steve is in a closed room with no windows, and he crosses from one side of the room to the other. Claire is outside the room. Infer that Claire does not know now that Steve has crossed the room.

The above problem is a somewhat dangerous benchmark for a theory, since it is trivially satisfied in a theory in which no one knows anything. To ensure that our theory is not trivial in this sense, we require that it support the following inferences as well:

II. Andrew is in his office and he does not see any cows there. Infer that he knows that there are no cows in his office.

III. Joanne, in the living room, does not know whether there are flowers on the dining room table. However, she knows that the top of the dining room table is visible from any point in the dining room. Infer that Joanne knows that she can find out whether there are flowers on the dining room table by going and looking.

IV. Fred has seen that Max has been with him all night, playing poker. Infer that Fred knows that Max was not five miles away, at the bank robbery.

V. Judy sees Sharon standing facing a bus, and looking at the bus. Infer that Judy knows that Sharon knows that there is a bus in front of her.

These problems illustrate various key aspects of perception. (I) shows how we can infer that an agent is ignorant of a fact from our knowledge of the physical limits of vision. Since Claire cannot see Steve inside the room, she cannot know what is happening in the room. (II) shows the gaining of knowledge from inferences based on both prior knowledge and perception. Andrew infers the absence of cows from the fact that, given what he knows about

I thank Leora Morgenstern and John Sterling for many helpful criticisms and suggestions. I also thank Ralph Grishman and Larry Manevitz for drawing my attention to the difficulties discussed in section 5. This work was supported by NSF grant #DCR-8603758.

the size of cows, the presence of a cow is physically incompatible with what he sees. (III) involves reasoning about possible future states of perception and knowledge. (IV) involves perception extended over an interval of time. (V) shows that one agent can infer the perceptions and knowledge of another agent by perceiving their physical situations and knowing their perceptual and inferential powers. The theory developed in this paper supports inferences of all these types.

The formal model developed in this paper is rich enough to support the statement and proof of close analogues of all the about inferences. Section 2 gives an informal characterization of the world described by the model. Section 3 describes the formal model, a variant on the well-known possible world semantics for knowledge [Hintikka, 71], [Moore, 80]. Also, in section 3, we show that our model implies a limited version of *negative introspection*: If an agent does not know some fact about the physical course of events, then he knows that he does not know it. Section 4 presents analogues of the above inferences, and sketches their proofs. Section 5 discusses some difficulties with the model, and section 6 presents our conclusions. The appendix gives the complete proof of an inference analogous to example (I).

It should be noted that our purposes are quite different from those of computer vision research: we therefore need a theory of perception at a very different level of description. There is no computer vision program that represents or uses the facts about vision needed to solve problems (I) through (V) above. For example, consider the fact that it is impossible to see what is happening outside a closed room from inside it, or the fact that it is hard to see from a brightly lit area into a dark area. Such facts are of little importance in computer vision, since they very rarely offer a significant constraint on the interpretation of a particular image. However, they are of great importance in reasoning qualitatively about how vision will augment knowledge in a given situation; one can use them to deduce that an agent will have to leave a closed room to know what is happening outside it, or that he will have to turn off the lights in his car to know what is on the highway. Conversely, our theory need not deal with the specific rules and algorithms necessary to interpret an actual image, because you do not generally have to interpret any images other than the one before you, which is tackled by the vision system. For example, the rule that a large change in detected light intensity may indicate a significant edge in an image is a key fact in image interpretation, but is of no value in inference about future or hypothetical state, since one cannot usually predict the light intensities to be detected.

A few limitations of our theory should be noted at the outset. First, it ignores learning through perceptions of conventional signs such as speech and writing. Perceptions are only used for their direct physical implications. Second, we assume that, under proper conditions, it can be known that a particular event has only limited physical effects. Third, we model only visual perception. We believe, however, that other forms of perception can be treated similarly. Finally, our models of vision and of physical causality are greatly simplified. We are interested here in the connection between perception and knowledge, and a more complex physical theory would just have added irrelevant complexity. We believe that the essential structure of our theory will carry over to more realistic physical models.

## 2. The Toy World

We formulate our theory within a highly simplified "toy" world, which we describe informally in this section. Many of the particular assumptions made below are obviously not valid in general; however, the basic structure does not depend strongly on these assumptions. In particular, the relations between knowledge and perception, which are our focus here, are largely independent of the particular physics of motion and vision assumed.

Our toy world is constructed as follows: There is a fixed set of physical objects scattered in space, moving about over time. Objects are rigid, maintaining a constant shape; however, their position may change continuously over time. At each time instant, each object occupies a *figure*; a connected regular[1] set of points. The places occupied by two objects at a single

1. A set is regular if it is closed, bounded, and equal to the closure of its interior.

time may not overlap. Besides their shapes, objects have time-invariant visible properties, such as their coloring. Objects may be also be characterized in terms of non-physical properties and relations, such as "being a Republican", or "being brothers."

There are no other physical restrictions on the world, beyond those mentioned above. A course of events is physically possible if and only if each object maintains a constant shape and properties and moves continuously, and no two objects ever overlap.

Some objects are *agents*. At each instant of time, an agent has a body of knowledge with the following properties:

A.1 Knowledge of axioms: All general axioms — axioms of predicate calculus, geometry, time, physics, knowledge and perception — are known.

A.2 Consequential closure: Any logical implication of the agent's knowledge is known.

A.3 Veridicality: All knowledge is true.

A.4 Positive introspection: If an agent knows a fact, he knows that he knows it.

A.5 Memory: If an agent knows a fact (with no time indexicals) at one time, he knows it at all later times.

A.6 Internal Clock: An agent always knows what time it is. (This assumption is not strictly necessary, but it makes things easier.)

An agent also has perceptions. A point is visible to an agent if it is not occluded from him by an object in between. (An obstacle $O$ occludes a point $P$ from an agent $A$ if any line segment from $A$ to $P$ intersects $O$ (Figure 1).) All objects are thus assumed to be opaque. An agent $A$ can see the visible properties of an object at a visible point. Moreover, $A$ can see the properties, not only of individual points but also of connected sets of points. Specifically, let $X$ be a connected set such that each point of $X$ is visible to $A$. If all of $X$ lies inside an object $O$ with visible properties $P1$, $P2$, ..., then $A$ sees that $X$ lies in some object with properties $P1$, $P2$, ... If $X$ lies in free space, then $A$ sees that $X$ lies entirely in free space.

Perception provides knowledge about the world, and it is the only source of information as to which of the many physically possible courses of events is actually happening. Specifically, we assume the following:

A.7 Anything that is perceived is known.

A.8 If a physical statement is physically possible, and it does not contradict any past or present perceptions, then it cannot be known to be false.

As we shall see in section 4, this model is rich enough to allow us to pose reasonable analogues of problems I through V above.

### 3. Formal Model

Our model of knowledge derives from Hintikka's (1971) possible worlds semantics for knowledge. Moore (1980) combined this with a temporal logic, by identifying epistemic possible worlds and temporal situations. We modify Moore's model in two ways. First, we adopt a continuous rather than a discrete model of time [McDermott, 82]. Second, and more importantly, we use two levels of possible worlds: *layouts* and *situations*. A *layout* is a timeless physical description of the instantaneous state of the world. A *situation* is a placement of a layout within a temporal structure and within a system of knowledge relations. Perceptions are associated with layouts; knowledge is associated with situations.

In our toy physics, an object is an atomic individual with a set of visual properties. A layout specifies the objects in the world and the figure occupied by each object. The figure occupied by an object $O$ in layout $L$ is denoted "place($O$,$L$)".

A *physical behavior* is an account of the physical layout over all time. Formally, it is a function from time to physical layouts with the following properties:

a. All layouts in the behavior have the same objects.

b. Objects have constant shape throughout the behavior. That is, for any object $O$, behavior $B$, and times $T1$ and $T2$, the two figures place($O$,scene($B$,$T1$)) and place($O$,scene($B$,$T2$)) are congruent.

c. Objects move continuously in time. That is, the function over time to figures $\lambda(T)$ place($O$,scene($B$,$T$)) is continuous, for any object $O$ and behavior $B$.

The function "scene($B$,$T$)" maps a behavior $B$ and a time $T$ to the layout of $B$ at time $T$. Time is real valued. The only relation that we will need on time is the total ordering, $T1$ is earlier than or equal to $T2$, written $T1 \le T2$.

We distinguish certain layouts and behaviors as *physically possible*. A layout is physically possible if no two objects overlap. A behavior is physically possible if each layout in the behavior is physically possible. We allow physically impossible layouts and behaviors as valid objects of thought; this simplifies the statement of the physical axioms. Note that conditions (a), (b), and (c) above apply even to physically impossible behaviors.

What the agent can see in a given physical layout is determined by the physical laws of vision. The perceptions of an agent $A$ in a layout $L$ fix all aspects of $L$ at points which are not occluded from $A$, and fix no aspects of $L$ at points which are occluded from $A$. We say that a layout $L1$ is visually compatible with $L$ with reference to $A$, written "v_compatible($A$,$L$,$L1$)" if $L1$ is consistent with everything that $A$ can perceive in $L$. V_compatibility is an equivalence relation over layouts.

Our theory of the scope and limits of vision is expressed in terms of the properties of the v_compatible relation. In our simple physics, $L2$ is v_compatible with $L1$ with respect to agent $A$ if both $L2$ and $L1$ are physically possible and the following condition holds: Let $X$ be a connected set of points, such that every point in $X$ is visible to $A$ in $L1$. Then each point of $X$ is visible to $A$ in $L2$. Moreover, if $X$ lies entirely in $A$ in $L1$, then $X$ lies entirely in $A$ in $L2$; if $X$ lies entirely in some object $O1$ in $L1$, then, in $L2$, $X$ lies entirely inside some object $O2$ with the same visible properties as $O1$; if $X$ lies entirely in free space in $L1$ then $X$ lies entirely in free space in $L2$.

A behavior $B1$ is visually compatible with behavior $B0$ for agent $A$ up to time $T$ if, as far as $A$ can see in $B0$ up to time $T$, the world could be going through $B1$. We write this relation, "bv_compatible($A$,$B0$,$B1$,$T$)". In our simple model of vision, we will assume that all such information comes through the layouts; that is, two physically possible behaviors are visually compatible if corresponding layouts are visually compatible.

$$\text{bv\_compatible}(A,B1,B2,TS) \Rightarrow$$
$$\forall_{T \le TS}\ \text{v\_compatible}(A,\text{situation}(B1,T),\text{situation}(B2,T))$$

Two states of the world may be identical in their physical layout and yet differ in other respects. To accommodate this, we define a *situation* as a state of the world, including the physical layout, the non-physical properties and relations of objects, and the knowledge states of agents. The knowledge states of agents, however, are not a component of the situation, but are encoded in accessibility relations between situations. Formally, we can define a situation as a triple $<L,\phi,X>$, where $L$ is a layout, $\phi$ is a function mapping each non-physical relation among objects to its extension as a set of tuples of objects, and $X$ is an atomic individual, which serves to distinguish situations that differ only in the associated knowledge states. (The problem here is that we want the option of having two different situations that are the same in all respects except the knowledge states. We accomplish this by giving them unequal values of $X$. The entity $X$ serves no other function.) The function "layout($S$)" maps a situation to its layout.

Non-physical properties of objects are made parts of situations rather than of layouts, in order to allow different agents to have different degrees of knowledge about them. In our system, we can allow Tom to know that all cows are large, but Sid not to know this. This would not be possible if we associated the non-physical property of being a cow with layouts.

The knowledge of an agent in a situation $S$ is represented by an accessibility relation between $S$ and other situations that are consistent with his knowledge. Let $A$ be an agent and let $S1$ and $S2$ be two situations. $S2$ is accessible from $S1$ relative to $A$, written "k($A$,$S1$,$S2$)", if as far $A$ knows in $S1$, the state of the world might be $S2$. We say that $A$ knows in $S1$ that $\phi$ is true if $\phi$ is true in all situations that are accessible to $A$ from $S1$.

Note that in both the visual compatibility relation and the knowledge accessibility relation, more information corresponds to a smaller extension of the relation. The more you know, the more variations in the world you can rule out as false, and therefore the fewer possible states of the world are consistent with your knowledge.

A *chronicle* is a function from time to situations. The function "situation($C$,$T$)" maps a chronicle $C$ and a time $T$ onto the situation in that chronicle at that time. We assume that each situation $S$ has a unique time in a unique chronicle, denoted "time($S$)" and "chronicle($S$)". (This restriction is not strictly necessary, but it simplifies descriptions without any loss of generality.) Each chronicle $C$ has associated a behavior, which is the progression of layouts of the scenes.

$$\text{behavior}(C) = \lambda(T)\ \text{layout}(\text{situation}(C,T))$$

If situations $S1$ and $S2$ are in the same chronicle, and the time of $S1$ is earlier than the time of $S2$, we say that $S1$ precedes $S2$.

$$\text{precedes}(S1,S2) \Leftrightarrow [\ \text{chronicle}(S1) = \text{chronicle}(S2) \wedge \text{time}(S1) \leq \text{time}(S2)\ ]$$

Knowledge about the past or the future is represented by combining the epistemic accessibility relation with the structure of chronicles. "Mark knows that the flower was in the vase at 9:00," is represented in the following form: "Let $S1$ be a situation accessible relative to Mark from the current situation s0. Let $C$ be the chronicle of $S1$. Let $S2$ be the situation of $C$ at 9:00. Then the place of the flower in $S2$ was inside the place of the vase in $S2$." (Figure 4).

$$\forall_{S1,S2}\ [\ \text{k}(\text{mark, s0}, S1) \wedge S2 = \text{situation}(\text{chronicle}(S1),\text{t900})\ ] \Rightarrow$$
$$\text{inside}(\text{place}(\text{flower1},S2),\ \text{place}(\text{vase1},S2))$$

We can achieve the properties of knowledge and perception enumerated above A.1 - A.8 by imposing the following requirements:

A.1, knowledge of the axioms, and A.2, consequential closure, follow immediately from the definition of knowledge, together with the axiom that all chronicles have a physically possible behavior.

$$\forall_C\ \text{phys\_poss}(\text{behavior}(C))$$

A.3, veridicality, is satisfied if the knowledge accessibility relation is reflexive.

$$\forall_{A,S}\ \text{k}(A,S,S)$$

A.4, positive introspection, follows if knowledge accessibility is transitive.

$$\forall_{A,S1,S2,S3}\ \text{k}(A,S1,S2) \wedge \text{k}(A,S2,S3) \Rightarrow \text{k}(A,S1,S3)$$

A.5, memory, holds given the following condition: Let situation $S1B$ be accessible from $S0B$, and let $S0A$ precede $S0B$ in the same chronicle. Then, since everything the agent knows in $S0A$, he also knows in $S0B$, and $S1B$ is consistent with everything he knows in $S0B$, there must be a scene $S1A$ in the chronicle of $S1B$ which is accessible from $S0A$. (Figure 5)

$$\forall_{A,S1B,S0B,S0A}\ [\ \text{k}(A,S0B,S1B) \wedge \text{precedes}(S0A,S0B)\ ] \Rightarrow$$
$$\exists_{S1A}\ [\ \text{k}(A,S0A,S1A) \wedge \text{precedes}(S1A,S1B)\ ]$$

A.6, internal clock, holds given the condition that, if $S1$ is accessible from $S0$, then the two situations occur at the same time.

$$\forall_{A,S0,S1}\ \mathrm{k}(A,S0,S1) \Rightarrow \mathrm{time}(S0) = \mathrm{time}(S1)$$

A.7, that perceptions are known, holds given the following: If $S1$ is knowledge accessible from $S0$, then the layout of $S1$ is visually compatible with the layout of $S0$. That is, for a situation $S1$ to be consistent with an agent's knowledge, the layout of $S1$ must be compatible with what the agent sees; conversely, if a layout is not compatible with what the agent sees, the agent knows that that cannot be the real layout.

$$\forall_{A,S0,S1}\ \mathrm{k}(A,S0,S1) \Rightarrow \mathrm{v\_compatible}(A,\mathrm{layout}(S0),\mathrm{layout}(S1))$$

A.8, that perception is the only source of knowledge of the course of events, holds given the following: Let $C0$ be the real chronicle. Let $B1$ be a behavior that is visually compatible with the behavior of $C0$ up to time $T$ relative to agent $A$; thus, as far as $A$ could have seen up to time $T$, $B1$ could be the real behavior. Then it is consistent with $A$'s knowledge that $B1$ actually was the real behavior; that is, there is a chronicle $C1$ whose situation in $T$ is knowledge accessible from the situation of $C0$ in $T$.

$$\begin{aligned}&\forall_{A,C0,B1,T}\ \mathrm{bv\_compatible}(A,\mathrm{behavior}(C0),B1,T) \Rightarrow \\ &\quad \exists_{C1}\ [\ \mathrm{k}(A,\mathrm{situation}(C0,T),\mathrm{situation}(C1,T)) \wedge B1 = \mathrm{behavior}(C1)\ ]\end{aligned}$$

Two general theorems of some interest follow from the axioms. The first is the analogue of axiom A.7 for perception over time: Facts perceived over time are known. Formally, if $S1$ is knowledge accessible from $S0$, then the behavior of the chronicle of $S0$ up $S0$ is visually compatible with the behavior of the chronicle of $S1$ up to $S1$.

$$\begin{aligned}&\forall_{A,S0,S1}\ \mathrm{k}(A,S0,S1) \Rightarrow \\ &\quad \mathrm{bv\_compatible}(A,\mathrm{behavior}(\mathrm{chronicle}(S0)),\ \mathrm{behavior}(\mathrm{chronicle}(S1)),\ \mathrm{time}(S0))\end{aligned}$$

The second theorem is a restricted form of *negative* introspection: If an agent does not know a fact about the course of events, then he knows that he doesn't know it. The formal statement is that, if a behavior is compatible with an agent's perceptions, then, from any knowledge accessible situation, there is a knowledge accessible chronicle exhibiting that behavior. That is, in all situations compatible with what the agent knows, there is a compatible chronicle in which the behavior occurs; i.e. the agent knows that he doesn't know that the behavior $B1$ didn't occur.

$$\begin{aligned}&\forall_{A,C0,B1,T,SM} \\ &\quad [\ \mathrm{bv\_compatible}(A,\mathrm{behavior}(C0),B1,T) \wedge \mathrm{k}(A,\mathrm{situation}(C0,T),SM)\ ] \Rightarrow \\ &\quad \exists_{C1}\ [\ \mathrm{k}(A,SM,\mathrm{situation}(C1,T)) \wedge \mathrm{bv\_compatible}(A,\mathrm{behavior}(C1),B1,T)\ ]\end{aligned}$$

The proofs of these two theorems are sketched in the appendix (Theorem 4.15 and 4.16).

## 4. Proofs

In this section we show how analogues to inferences (I) through (V) can be formulated in our model, and we will sketch how they can be proven from the above axioms of perception and knowledge, together with suitable axioms of geometry and physics. In our discussion we will show some of the logical formulas that arise in these proofs, including the logical statement of the conclusion of the proof. We will not here define all the non-logical symbols introduced. Most of them are defined in the appendix; the rest can be easily deciphered using context and English translation. Quantified variables are indicated with italicized capital letters; free variables are assumed to quantified over the entire formula. The complete proof of example (I) may be found in the appendix. This complete proof uses a standard first-order sorted logic, with equality and function symbols.

I. Claire is on one side of a wall, for some interval of time. On the other side of the wall, occluded from Claire, is an object omystery. The object lies strictly within some larger region, which is entirely occluded from Claire. (Figure 2). Over a certain time interval i0, Claire stays motionless, the object stays within its envelope, and no other object ever intersects the envelope. We wish to prove that there is no way for Claire to know whether the object is motionless or whether it is moving around within its envelope, since either are equally compatible with the motions of the objects that Claire does see.[2] This conclusion can be formalized as follows: At the end of i0, there is one knowledge accessible situation that follows on a chronicle in which the object is motionless; there is another knowledge accessible situation that follows on a chronicle in which the object is in motion.

$$[\, \exists_{C1} \text{k(aclaire,s0z, situation}(C1\text{,end(i0)))} \wedge \text{motionless(omystery,behavior}(C1)\text{,i0)} \,] \wedge$$
$$[\, \exists_{C2} \text{k(aclaire,s0z,situation}(C2\text{,end(i0)))} \wedge \neg\text{motionless(omystery,behavior}(C2)\text{,i0)} \,]$$

To prove this, we construct two particular behaviors. In the first, every object moves just as it does in the real world except that omystery stays motionless throughout i0. In the second every object moves just as it does in the real world except that omystery moves continuously within xenvelope throughout i0. We show that both of these are physically possible, since no other object comes within the envelope, by hypothesis, and so no other object interacts with omystery. Both compatible with Claire's perceptions, since the identical objects are visible to Claire in the identical places. Hence, by axiom A.8, Claire cannot know of either of them that it did not occur.

II. Andrew knows that he is in a closed square room 20 feet by 20 feet. He knows that all cows contain a circular disk 2 feet in diameter. Andrew is in the center of the room. Every other object in the room lies entirely within a foot of the wall. Thus, from what Andrew knows of the room of cows, and from what he can see of the other objects, he can infer that no cow could fit into the occluded parts of the room. We wish to prove that Andrew knows that there is no cow in the room.

This conclusion can be formalized as the statement that in any knowledge accessible world, any cow is outside the room.

$$[\, \text{k(andrew},S0,S1) \wedge \text{true\_in}(S1,\text{cow}(OCOW)) \,] \Rightarrow$$
$$\neg\text{sub\_place(place}(OCOW,\text{layout}(S1)),\ \text{place(inside(oroom1)},S1)) \,]$$

Note that we treat "being a cow" as a contingent property; otherwise, we would be committing to supposing that everyone would always know whether a given object was a cow. The term "inside(oroom1)" here designates a *pseudo-object;* that is, a point set that moves around with the room [Davis, 86b].

Andrew's knowledge that all cows contain a disk two feet in diameter can be formalized as the statement that, in all knowledge accessible worlds, all cows have that property.

$$[\, \text{k(andrew},S0,S1) \text{ and true\_in}(S1,\text{cow}(OCOW)) \,] \Rightarrow$$
$$\exists_X \text{ disk}(X,\text{two\_feet}) \wedge \text{sub\_place}(X,\text{place}(OCOW,S1))$$

This assumption does not contradict axiom A.8, that all information about the course of events is gathered from perception, if we assume that being a cow is not an aspect of the physical course of events and does not enter into layouts, but only into situations. We will discuss this point further in section 5.

We are given that Andrew knows lower bounds on the size of cows, that he knows the size and shape of the room, and that all objects in the room other than himself lie within a foot of the walls of the room. By a geometrical argument, then, in any layout compatible

2 The particular geometry chosen here is not, by any means, the only one that will allow the argument to go through; however, it simplifies the geometrical analysis. The example of a closed room is too strong: in our theory an actor outside a closed room can never find out anything about what happens inside the room.

with what he sees, the entire center of the room is clear of objects. (Note that, since the objects partially occlude the walls, Andrew cannot at the moment see that they are all within a foot of the walls; that is, there are compatible layouts in which the walls are further back.) By axiom A.7, that perceptions are known, it follows that in any knowledge accessible situation, the entire center of the room is clear. But Andrew knows the structure of the room; that is, in all knowledge accessible situations, the walls of the room have their true shape. Therefore, in all knowledge accessible situations, all visible objects lie within a foot of the wall. Therefore, by another geometrical argument, in every accessible world, the room does not contain any object containing a disk of two foot diameter, since such an object would either overlap the walls or the empty center space. But in all accessible situations, any cow is contains a two-foot diameter disk. Hence, in all accessible situations, there is no cow in the room.

III. Joanne is inside some enclosure with a cat and with a U-shaped room. Joanne knows the shape of the room. We wish to prove that Joanne knows that, if she, the cat, and the room are the only things within the enclosure, and she goes into the room, then she will know whether the cat is also in the living room (Figure 3).

This conclusion can be formalized as follows: Let $S0$ be the real situation. Let $S1$ be any situation knowledge accessible from $S0$ such that, in $S1$, Joanne, the cat, and the room are the only objects in the enclosure. Let $S2$ be a situation later than $S1$ in which Joanne is in the room. Then one of two things are true: Either, in all situations accessible from $S2$, the cat is in the room, or, in all situations accessible from $S2$, the cat is not in the room.

$$
\begin{array}{l}
[\ \mathrm{k(ajoanne},S0,S1) \wedge \\
\quad [\ \forall_O \ \mathrm{sub\_place(place}(O,\mathrm{layout}(S1)),\mathrm{place(inside(oenclosure)},\mathrm{layout}(S1))) \Rightarrow \\
\qquad [\ O=\mathrm{ajoanne} \vee O=\mathrm{ocat} \vee O=\mathrm{oroom}\ ]\ ] \wedge \\
\quad \mathrm{precedes}(S1,S2) \wedge \\
\quad \mathrm{sub\_place(place(ajoanne},\mathrm{layout}(S2)),\mathrm{place(inside(oroom)},\mathrm{layout}(S2))) \\
] \Rightarrow \\
[\ [\ \forall_{S3}\ [\mathrm{k(ajoanne},S2,S3) \Rightarrow \\
\qquad \mathrm{sub\_place(place(ocat},\mathrm{layout}(S3)),\mathrm{place(inside(oroom)},\mathrm{layout}(S3)))\ ] \vee \\
\quad [\ \forall_{S3}\ [\mathrm{k(ajoanne},S2,S3) \Rightarrow \\
\qquad \neg\mathrm{sub\_place(place(ocat},\mathrm{layout}(S3)),\mathrm{place(inside(oroom)},\mathrm{layout}(S3)))\ ]\ ]
\end{array}
$$

Proof: If Joanne, the cat, and the room are the only objects inside the enclosure in situation $S1$, they must be the only objects inside the enclosure in situation $S2$, since, by physical laws, nothing can go through a closed barrier. By a geometrical argument from the shape of the room we can show that, (i) if Joanne is the only object in the room, then she can see the entire inside area of the room as well as the inner walls of the room, and thus determine that she is inside the room, and that nothing else is; (ii) if Joanne and the cat are both inside the room, then the cat will be visible and enough of the inner walls of the room will be visible that Joanne will be able to see that she is inside the room. (Keep in mind that Joanne not only has to see the cat at a place that is in fact inside the room; she also has to see enough of the room to know that the place of the cat is inside the room.) Either case (i) or case (ii) must hold. Using axiom A.7, that what is seen is known, we deduce that, in case (i) Joanne will know that the cat is not inside the room, while in case (ii), Joanne will know that the cat is inside the room.

Two limitations of our theory should be noted in connection with this example. The first is that we cannot express the fact that Joanne can go into the room if she chooses to do so, since we have no planning vocabulary. The second is that, due to the extremely strong requirement that facts about the course of events can only be known through perception, there is no way that Joanne can ever find out that there are no other objects in the enclosure. The other objects could just have hidden from her up to now, and be ready to spring out and surround her when she tries to look for the cat. We will discuss this further in section 5.

IV. Fred knows that Max has a diameter less than seven feet. During the time interval i0, Max was always within ten feet of Fred, and Fred always has an unoccluded view of him. We wish to prove Fred knows that Max was not one hundred feet away during any part of i0.

Formally, let $S1$ be a knowledge accessible situation, let $C1$ be the chronicle of $S1$, and let $T$ be a time in i0. Then in the situation of $C1$ at $T$, no point of Max is one hundred feet from any point of Fred.

$$[\ \mathrm{k(afred}, S0, S1) \wedge C1 = \mathrm{chronicle}(S1) \wedge \mathrm{time\_in}(T, \mathrm{i0}) \wedge LT = \mathrm{layout(situation}(C1, T)) \wedge \mathrm{point\_in}(PF, \mathrm{place(afred}, LT)) \wedge \mathrm{point\_in}(PM, \mathrm{place(amax}, LT))\ ] \Rightarrow \mathrm{distance}(PF, PM) < 100 * \mathrm{foot}.$$

Proof: By the theorem of perception over time, any chronicle containing a knowledge accessible situation must have a behavior compatible with the agent's perceptions up to that time. But Fred perceives throughout the time interval i0 that some point of Max is within ten feet of him, so any knowledge accessible chronicle must have some point of Max within ten feet of Fred at all times. Moreover, Fred knows that Max has a diameter of less than seven feet, so that, in any knowledge accessible chronicle, any point of Max is less than seven feet from any other point of Max. By the triangle inequality, then, in any knowledge accessible chronicle during i0, any point of Max is less than seventeen feet from Fred.

V. Judy, Sharon, and the bus are convex objects. No other object intersects the convex hull of the three of them. We assume that Sharon has a unique set of visible properties that allow her to be distinguished from anyone else. Both Judy and Sharon recognize busses; that is busses have visual properties that Judy and Sharon know attach only to them. Due to the geometry of the situation, Judy can see Sharon, the bus, and the space between them. We wish to show that Judy knows that Sharon knows that there is a bus in front of her.

The formal statement of the conclusion is as follows: Let $S0$ be the real situation, let $S1$ be any situation accessible from $S0$ through Judy's knowledge, and let $S2$ be any situation accessible from $S1$ through Sharon's knowledge. Then in $S2$ there is a bus in front of Sharon.

$$[\ \mathrm{k(ajudy}, S0, S1) \wedge \mathrm{k(asharon}, S1, S2)\ ] \Rightarrow [\ \exists_{OBUS}\ \mathrm{true\_in}(S2, \mathrm{bus}(OBUS)) \wedge \mathrm{in\_front(place}(OBUS, \mathrm{layout}(S2)), \mathrm{place(asharon, layout}(S2)))\ ]$$

The statement that Judy recognizes busses can be formalized as the statement, "In any situation knowledge accessible to Judy, any object with the same visual properties as some bus is also a bus."

$$[\ \mathrm{k(ajudy}, S0, S1) \wedge \mathrm{same\_vprops}(O, OBUS) \wedge \mathrm{true\_in}(S1, \mathrm{bus}(OBUS))\ ] \Rightarrow \mathrm{true\_in}(S1, \mathrm{bus}(O))$$

Proof: Since Judy's view of Sharon, the bus, and the space between them is unblocked, any layout visually compatible with her perceptions will have to have Sharon in Sharon's place, some object indistinguishable from a bus in the bus's place, and the space between unblocked. Since Judy recognizes busses, the object indistinguishable from a bus must, indeed, be a bus. By A.7, that knowledge accessible states have visually compatible layouts, this same geometrical constraint must apply in any situation $S1$ knowledge accessible to Judy. Thus, Sharon's view of the bus in the layout of $S1$ is unblocked; therefore, in any layout visually compatible with $S1$ to Sharon, there must be a bus in its place. Therefore, applying axiom A.7 again, in $S1$ Sharon must know that there is a bus in front of her.

## 5. The Problem with Inferring Ignorance

Our proof of inference (I), that Claire cannot know whether the mystery object is moving or still, rests on axiom A.8, which limits what an agent can know about the course of events to what he can perceive. Specifically, the axiom states that, if a course of events is possible (we have above specified "physically possible", but broader classes of possibility can

be considered) and it is compatible with the agent's perceptions over time, then the agent cannot know that this is not the course of events that took place. Clearly, some such axiom is needed if we are to deduce the ignorance of agents from the limits of their perceptions.

Unfortunately, axiom A.8 is so strong that it rules out many plausible states of knowledge. This problem is particularly exacerbated by the weakness of our physics, which allows all kinds of courses of events as possible. For example, when I first formulated example (III), I wanted to specify that Joanne knew that she, the cat, and the room were the only objects in the enclosure. However, axiom A.8 makes this impossible. Whatever Joanne has done or seen in the past, her perceptions were necessarily consistent with the course of events in which there is a swarm of small mobile objects that have until now stayed hidden from her on the opposite sides of the room. Therefore by axiom A.8, she does not know that this swarm does not exist. If so, then there is no way that she can predict that she will ever see either the cat or the room again, since the swarm could at any moment come and surround her and block her entire vision.

Another example: It would be nice to augment the physics by allowing certain objects to be specified as immobile. Such a physics would be substantially more realistic, and would simplify many inferences. For example, as things stand now, in example III, Joanne must know that she will see enough of the inside of the room to know that she is inside; this would not be necessary if Joanne could know that the room is fixed. It is, in fact, easy enough to change the physical axioms to accommodate this; all that is required is to add the requirement that immobile objects do not change position in any physically possible behavior. However, under Axiom A.8, it is necessary either that everyone always know which objects are immobile, or that no one ever knows which objects are immobile. If we define immobility as a situation-independent property of objects, then each agent must always know the mobility of each object. If we allow an object to be immobile in one behavior and mobile in another, then there is no way to distinguish visually between the case where the object is actually immobile, and can be trusted to stay in the same place, and the case where it is mobile in principle but has simply chosen to stay in the same place up till now. Since they are not visually distinguishable, an agent cannot know which is the case. Of course, we could eliminate this particular problem by properly enriching the physics with a rule such as "If a mobile object is hit, it will necessarily move," but how are we to know that the new physics will not suffer from the same problem as regards some other property? In general, given a physical theory and a property in that theory, it is difficult to tell whether it is consistent with axiom A.8 that agents should sometimes but not always know the value of that property. Moreover, it is objectionable to have our epistemic theory dictate the form of our physical theory.

Thus, axiom A.8 in its current form rules out problem formulations that seem entirely natural. Moreover, it does so in quite subtle ways. It took me personally a while to discover the inconsistency in the original formulation of example III; it would probably not be feasible to do so automatically. Even worse, it becomes quite difficult to convince oneself that a problem formulation is consistent. Each time you specify that an agent knows something, you have to ask yourself whether he could possibly know it, given axiom A.8.

For example, we specify in example IV that Fred knows that Max has a diameter of less than seven feet. Before going further with the problem, we have to check whether this is possible; can Fred possibly know this? As it happens, the answer is that he can, under the proper circumstances. Two things are required: Fred must know that this object he is seeing has diameter less than seven feet, and he must know that it is Max. Fred can know the diameter if he has walked around Max and seen all sides of him. Moreover, he must know that he had seen all sides of Max. This is a non-trivial requirement. For example, as far as I can tell, there is no way that anyone in this theory can ever know that an object is circular. Even if you walk around it, you don't know whether you have seen all sides of it, or whether it has been rotating with you, like the moon, and keeping the irregularities on its backside hidden from you. For Fred to know that the object he sees is Max, we must posit that Max is visually distinguished from every other object; if Max had a Doppelganger, then there is no way

for Fred to know that he is not seeing the Doppelganger. As this analysis suggests, in this theory justifying a problem statement by constructing scenarios in which the agents have the stated knowledge can be as much work as solving the problem.

The system does allow great freedom in specifying knowledge about situation dependent properties: if a property or relation is not an element of the layout, axiom A.8 does not apply, and hence does not rule knowing about it. Thus, we tend to define as much as we can as situation-dependent properties; in the above examples, we have treated "being a cow" and "being a bus" as situation dependent. Axiom A.8 does not prevent us from knowing that cows are larger than breadboxes, because that fact does not constrain layouts in any way. It is consistent with any layout; it affects only how the situation categorizes the objects of the layouts as cows or non-cows. This freedom, however, suffers from two critical limitations. The first reason is that one can, so to speak, tie down a situation dependent property to a physical layout only at one point; if you try to tie it down at two points, you run into trouble. For example, an agent can know that a cow is larger than a breadbox; alternatively an agent can be able to recognize a cow; that is, to know that anything of specified visible properties is a cow. However, an agent cannot do both. If he did, he would know that anything with those properties is larger than a breadbox. This is not allowed by axiom A.8, since there can perfectly well be layouts in which there are objects smaller than breadboxes with the given properties.

The second limitation, which is related to the first, is that what we are allowing agents to know is actually trivial, particular in the context of our theory, which so strongly emphasizes knowledge of physical facts. Our theory allows an agent to know that there are no unicorns, if "unicorn" is taken as a situation dependent property. However, it does not allow the agent to know that there are no horse-shaped objects with a single horn, which is what he really wants to know when push comes to shove. We can fix this by containing layouts including such objects as impossible. However, this has the unfortunate consequence that now all agents must know that there are no unicorns; it is no longer possible for one agent to know it and for another agent to be ignorant of it.

One possible fix is to change A.8 to read as follows: "If a behavior $B$ is compatible with an agent's perceptions, and *if the agent does not know that* B *is impossible*, then the agent cannot know that $B$ has not occurred." This formulation is clearly in some ways more reasonable than the formulation that we have put forward, since it allows agents to have different degrees of knowledge about what is possible. If one agent knows more physics than another, he will be able to make tighter predictions about what will happen. However, from the point of view of developing a useful theory for inference, this axiom throws out the baby with the bathwater. The point of the entire exercise was to be able to deduce that the agent was ignorant of one fact, purely by knowing about what he has seen, without knowing that he is ignorant of some other fact. By contrast, to use this axiom to deduce that the agent does not know that $B$ does not occur, we must be first be able to deduce that the agent does not know that $B$ is not impossible. The latter deduction involves the same kinds of problems as the first. The only case where it is easy to deduce that the agent does not know that $B$ is impossible is when $B$ is, in fact, possible; but this is just the case covered by our original formulation.

Another possible fix would be to change the axiom to read. "If behavior $B$ is compatible with the agent's perceptions between times $T1$ and $T2$, and the agent does not know in $T1$ that $B$ does not occur, then he cannot know in $T2$ that $B$ does not occur." In other words, perception limits what an agent can find out but not what he can know. (Other knowledge would be innate.) Unfortunately, this axiom turns out to be too restricted to be applied in most cases. Consider example I, where we are trying to deduce that Claire does not know whether the object is moving or is standing still. The inference we would want to make is as follows: If Claire does not know at the beginning of i0 whether the object will move or not, then, since what Claire can see is physically consistent either with the object moving or with its standing still, Claire will not know at the end of i0 whether the object has

moved or not. Unfortunately, this does not follow, since Claire might know any number of things that relate what she can see to what the object does. For instance, it is consistent both with the axiom and with the problem formulation that Claire should know the statement, "The object will move during i0 if and only if the cat (assumed to be visible) moves during i0." If we allow Claire to know facts like this innately, then the inference is not legitimate; it is hard to disallow facts like this in a general rule, and clumsy to do so in the particular problem formulation.

Our problem here is to distinguish between facts like, "There is nothing that looks like a horse with a horn," which we want to allow, but not compel, agents to know, and facts like, "The object will move during i0 if and only if the cat moves during i0," which we wish to compel agents not to know. Intuitively, the difference is that the first statement is a general law, which can be inferred inductively from the failure to encounter any unicorns, while the second is neither a general law nor an instance of any general law. However, it does not appear that a possible worlds theory supports this kind of distinction; we may have to move to a syntactic theory of knowledge to accommodate it.

The ultimate problem here is that the only kind of knowledge acquisition we are considering is deductive inference through perception, and this is simply not adequate for many types of knowledge. We have, in fact, fallen into Berkeley's position, that nothing can be known except the pure perceptions, except that we have arbitrarily given agents innate knowledge of what is physically possible and impossible. In reality, if a person explores a territory and sees nothing, he is entitled to believe that there is nothing there, and to discount the possibility that something is cunningly hiding from him. If he is right, we can call this knowledge. Our theory of knowledge, like most such theories, approximates rationality in terms of the axiom of consequential closure, the assumption that an agent can make all logical deductions. It has often been pointed out that this axiom is too strong (e.g. [Konolige, 86], [Levesque, 84]); it is much less often noted that it is also too weak, and that a complete account of knowledge must contain a description of plausible reasoning.

Another discomfitting feature of our theory is more fundamental than the particular axiom A.8. Our theory provides three ways to define properties of objects; they can be situation-dependent, layout-dependent, or absolute. The criteria for choosing one or another are not always intuitive and the consequences of one choice or another are not always obvious. It often requires a delicate balancing act to come up with a structure that is powerful enough to support desired conclusions, but not so powerful as to support preposterous conclusions. Even if a property is conceived as a necessary property of an object, it may make a difference whether it is axiomatized as holding in all situations, holding in all layouts, or simply holding absolutely.

**6. Previous Work**

Little work has been done in AI on reasoning about perception. ATTEND, the focussing of a sensory organ, was a primitive act in conceptual dependency [Schank, 75] and was causally connected to MBUILD, the performance of a mental act; but the logic of these acts was not developed in detail. A number of theories, such as [Appelt, 82], [Allen and Perrault, 80] and [Morgenstern, 87], have studied the acquisition of knowledge through communication, but these have not looked at direct sensation. The definition of a perception as a set of physical layouts was put forward in [Davis, 84]. Hintikka (1969) used a modal theory of perceptions to eliminate the need for sense-data as ontological primitives. The situation semantics of [Barwise and Perry, 82] studied the logical structure of sensory verbs in detail, but did not relate it to knowledge acquisition or to physical constraints. Reiter and Mackworth (1987) give a formal account of the relation between an image and a physical situation.

## 7. Conclusions and Future Work

We have given a formalism in which a few basic problems relating perception and knowledge can be stated and solved. A few features of our approach should be pointed out:

There are two significant technical innovations in this work. The first is the introduction of the concept of a physical layout, which specifies just the physical state of the world, and the description of perception in terms of layouts. The second is the axiom A.8, which limits an agent's knowledge of contingent physical facts of certain types to that which can be deduced from his perceptions, together with physical laws. This axiom, however, places constraints on an agent's knowledge that are unacceptably strong in many cases. In most practical problems, the positive inference, "Since $A$ sees $\phi$, $A$ knows $\phi$," is more important that the negative inference, "Since $A$ doesn't see $\phi$, $A$ cannot know $\phi$;" hence, it may be best to drop axiom A.8 or to restrict its scope.

The theory of the connection between perception and knowledge is largely independent of the physical theory. In particular, axioms A.1 through A.7 may be used together with any set of physical laws, and with any laws delimiting the powers of vision in terms of bv_compatibility. Axiom A.8 cam also be used with any definition of physical laws and laws of vision, but it may put strong limits on the range of agents' knowledge of physical properties; that is, it may require that either all agents always know the value of a particular physical property or that no agent ever know it.

We used a possible-worlds theory of knowledge, as in [Moore, 80] rather than a syntactic theory of knowledge, as in [Konolige, 86] for a couple of reasons. First, there is a technical difficulty with syntactic theories. In a possible worlds theory, we can easily say that a person knows facts about each connected subset of the plane. However, this is an uncountable set; hence, a syntactic theory must either be restricted to sets definable in first-order formulas, or must start with an uncountable alphabet of symbols. Second, the use of physically possible layouts and behaviors is basic to this theory. These concepts go well with possible worlds semantics; their use in a syntactic theory seems forced. Finally, we simply had no need for the power of a syntactic theory. However, we plan to study reformulations of this theory in which "know" is a syntactic operator.

We plan to extend this theory in several ways. We hope to incorporate more realistic theories of vision and of physics and to add other modes of perception; to study the connection between perception and prior knowledge in greater depth, in particular the issues involved in recognizing specific objects; to integrate our theory with a theory of communication through writing and speech and with a theory of plans [Morgenstern, 87]; and to apply our theory to the problem of forming cognitive maps [Davis, 86a].

## 8. Appendix: Formal Proof

In this appendix, we give a formal proof of our first sample inference. We use a first-order sorted logic with equality. Since our primary interest here is in perception and knowledge, rather than physics and geometry, we have taken many fairly high-level geometrical and physical theorems as axioms. It would be possible to derive these from basic axioms in well-known ways; however, the resultant proofs would be unacceptably long, and largely irrelevant to our purposes. We have explicitly avoided the use of set theoretic notation, in order to illustrate that problems of this kind require only very shallow sets, and in order to simplify the structure of sorts needed. If set theory were added, it would be possible to replace some of the non-logical symbols by standard set-theoretic relations, and to eliminate some of the axioms.

The sorts of variables and constants in our logic will be indicated by their first letter, using the key in table 1:

| First Letter | Sort |
|---|---|
| $A$ | Agent |
| $B$ | Behavior |
| $C$ | Chronicle |
| $I$ | Closed Interval of Time |
| $L$ | Layout |
| $O$ | Object |
| $P$ | Geometric Point |
| $S$ | Situation |
| $T$ | Instant of Time |
| $X$ | Point set |

Table 1: Logical Sorts

"Agent" is a sub-sort of "object"; hence, any position that can accept an argument of sort "object" can accept an argument of sort "agent", though not vice versa.

We will indicate variables by symbols beginning with upper-case letters; symbols beginning with lowercase letters will be non-logical (predicate, function, or constant) symbols. Free variables are taken to be universally quantified with a scope of the entire formula.

Definitions are sometimes followed by logical formulas. Sometimes these are equivalent in meaning; sometimes, they capture only part of the meaning. Whatever is in a definition that is not expressed in a following logical formula is given only as an informal explanation. The proofs rest only on the logical formulas given for the definitions and axioms.

## 8.1. Geometry

**Definition 1.1:** The predicate "point_in($P$,$X$)" holds if $P$ is a point in $X$.

**Axiom 1.2:** A point set is determined by the points in it.

$$X1 = X2 \Leftrightarrow \forall_P \text{ point\_in}(P,X1) \equiv \text{point\_in}(P,X2)$$

**Definition 1.3:** The predicate "sub_place($X1$,$X2$)" if $X1$ is a subset of $X2$.

$$\text{sub\_place}(X1,X2) \Leftrightarrow \forall_P \text{ point\_in}(P,X1) \Rightarrow \text{point\_in}(P,X2)$$

**Lemma 1.4:** The sub_place relation is a partial ordering.

$$\text{sub\_place}(X1,X1).$$
$$[\text{ sub\_place}(X1,X2) \wedge \text{sub\_place}(X2,X1) ] \Rightarrow X1 = X2.$$
$$[\text{ sub\_place}(X1,X2) \wedge \text{sub\_place}(X2,X3) ] \Rightarrow \text{sub\_place}(X1,X3)$$

**Proof:** Immediate from definition 1.3 and axiom 1.2.

**Definition 1.5:** The predicate "intersect($X1$,$X2$)" holds if the two point sets intersect.

$$\text{intersect}(X1,X2) \Leftrightarrow \exists_P \text{ point\_in}(P,X1) \wedge \text{point\_in}(P,X2)$$

**Definition 1.6:** The function "interior($X$)" maps a point set $X$ to its interior. (Interior here is used in the topological sense; the interior of $X$ is $X$ minus its boundary.)

**Axiom 1.7:** If $X1$ is a subset of $X2$, then the interior of $X1$ is a subset of the interior of $X2$.

$$\text{sub\_place}(X1,X2) \Rightarrow \text{sub\_place}(\text{interior}(X1),\text{interior}(X2))$$

**Definition 1.8:** The predicate "overlap($X1$,$X2$)" holds the interiors of $X1$ and $X2$ intersect.

$$\text{overlap}(X1,X2) \Leftrightarrow \text{intersect}(\text{interior}(X1),\text{interior}(X2))$$

**Lemma 1.9:** If $X1$ overlaps $XA$ and $XA$ is a subset of $XB$, then $X1$ overlaps $XB$.

$$[\ \text{overlap}(X1,XA) \wedge \text{sub\_place}(XA,XB)\ ] \Rightarrow \text{overlap}(X1,XB)$$

**Proof:** Immediate from Definitions 1.3, 1.5, and 1.8 and axiom 1.7.

**Definition 1.10:** The predicate "connected($X$)" holds if $X$ is a connected set of points.

**Definition 1.11:** The predicate "strict_inside($X1$,$X2$)" means that $X1$ is a subset of $X2$ and, moreover, their boundaries are disjoint.

$$\text{strict\_inside}(X1,X2) \Rightarrow \text{sub\_place}(X1,X2)$$

**Definition 1.12:** The function "line_seg($P1$,$P2$)" maps two points $P1$ and $P2$ onto the line segment connecting them (a point set).

**Axiom 1.13:** "Line_seg($P1$,$P2$)" is a symmetric function.

$$\text{line\_seg}(P1,P2) = \text{line\_seg}(P2,P1)$$

**Axiom 1.14:** $P1$ lies on the line segment from $P1$ to $P2$.

$$\text{point\_in}(P1,\text{line\_seg}(P1,P2))$$

**Definition 1.15:** The predicate "blocked($PA$,$X$,$PB$)" means that point set $X$ blocks the view of point $PA$ from point $PB$. (The order corresponds to the physical layout: $X$ comes between $PA$ and $PB$.). That is, the line from $PA$ to $PB$ intersects the interior of $X$.

$$\text{blocked}(PA,X,PB) \Leftrightarrow \text{intersect}(\text{interior}(X),\ \text{line\_seg}(PA,PB))$$

We will overload the predicate "blocked" to take point-set arguments either in its second and third arguments or in all three arguments.

**Definition 1.16:** The predicate "blocked($P$,$XB$,$XC$)" means that $P$ is blocked by $XB$ every point in $XC$. The predicate "blocked($XA$,$XB$,$XC$)" means that each point of $XA$ is blocked by $XB$ from any point in $XC$.

$$\text{blocked}(P,XB,XC) \Leftrightarrow \forall_{PC}\ [\ \text{point\_in}(PC,XC) \Rightarrow \text{blocked}(P,XB,PC)\ ]$$

$$\text{blocked}(XA,XB,XC) \Leftrightarrow \forall_{PA}\ [\text{point\_in}(PA,XA) \Rightarrow \text{blocked}(PA,XB,XC)\ ]$$

**Lemma 1.17:** If $XA$ is blocked by $XB$ from $XC$, then any subplace of $XA$ is blocked from any subplace of $XC$.

$$[\ \text{blocked}(XA,XB,XC) \wedge \text{subplace}(XA1,XA) \wedge \text{sub\_place}(XC1,XC)\ ] \Rightarrow$$
$$\text{blocked}(XA1,XB,XC1)$$

**Definition 1.18:** The predicate "unblocked($XA$,$XB$,$XC$)" holds if no point in $XA$ is blocked by $XB$ from $XC$.

$$\text{unblocked}(XA,XB,XC) \Leftrightarrow$$
$$\forall_{PA}\ [\ \text{point\_in}(PA,XA) \Rightarrow \neg\text{blocked}(PA,XB,XC)\ ]$$

## 8.2. Time and Motion

**Definition 2.1:** The predicate "$T1 \leq T2$" (written infix) holds if time instant $T1$ is earlier or equal to $T2$.

**Axiom 2.2:** Time instants are totally ordered.

$$T1 \le T2 \vee T2 \le T1$$
$$[T1 \le T2 \wedge T2 \le T1] \Rightarrow T1 = T2$$
$$[T1 \le T2 \wedge T2 \le T3] \Rightarrow T1 \le T3$$

**Definition 2.3:** The function "start($I$)" maps a time interval $I$ to its starting time instant.

**Definition 2.4:** The function "end($I$)" maps a time interval $I$ to its ending time instant.

**Definition 2.5:** The predicate "time_in($T$,$I$)" holds if the time instant $T$ is part of the time interval $I$.

$$\text{time\_in}(T,I) \Leftrightarrow \text{start}(I) \le T \le \text{end}(I)$$

**Axiom 2.6:** Any two ordered unequal time points determine an interval.

$$[\,T1 \le T2 \wedge T1 \ne T2] \Rightarrow \exists_I\, [T1 = \text{start}(I) \wedge T2 = \text{end}(I)\,]$$

**Definition 2.7:** The predicate "subinterval($I1$,$I2$)" holds if interval $I1$ is a subset of interval $I2$.

$$\text{subinterval}(I1,I2) \Rightarrow [\,\forall_T\ \text{time\_in}(T,I1) \Rightarrow \text{time\_in}(T,I2)\,]$$

**Definition 2.8:** The function "scene($B$,$T$)" maps a behavior $B$ and a time $T$ onto the layout of $B$ at time $T$.

**Axiom 2.9:** Two layouts are equal just if they have the same objects, and they assign them to the same places. (We do not consider rotations in place to make a difference).

$$\begin{aligned} &L1 = L2 \Leftrightarrow \\ &\forall_O\, [\,\text{object\_of}(O,L1) \Leftrightarrow \text{object\_of}(O,L2)\,] \wedge \\ &\quad [\,\text{object\_of}(O,L1) \Rightarrow \text{place}(O,L1) = \text{place}(O,L2)\,] \end{aligned}$$

**Axiom 2.10:** Two behaviors are equal just if corresponding layouts are equal.

$$B1 = B2 \Leftrightarrow \forall_T\ \text{layout}(B1,T) = \text{layout}(B2,T)$$

**Definition 2.11:** The function "chronicle($S$)" maps a situation $S$ onto the chronicle containing it.

**Definition 2.12:** The function "time($S$)" maps a situation $S$ onto the time when it occurs.

**Definition 2.13:** The function "situation($C$,$T$)" maps a chronicle $C$ and a time instant $T$ into the situation of $C$ at time $T$.

$$\text{situation}(\text{chronicle}(S),\ \text{time}(S)) = S\,.$$

**Definition 2.14:** Situation $S1$ precedes $S2$ if they belong to the same chronicle and $S1$ occurs earlier.

$$\text{precedes}(S1,S2) \Leftrightarrow \text{chronicle}(S1) = \text{chronicle}(S2) \wedge \text{time}(S1) \le \text{time}(S2)$$

**Definition 2.15:** The function "layout($S$)" maps a situation $S$ into the physical layout present in $S$.

**Definition 2.16:** The function "behavior($C$)" maps a chronicle $C$ onto its behavior.

**Axiom 2.17:** A chronicle at a given time has the same layout whether you go through the behavior or through the situation.

$$\text{scene}(\text{behavior}(C),\ T) = \text{layout}(\text{situation}(C,T))$$

Note: The next several axioms could all be proved in greater generality from fundamental definition of a layouts as a functions from a set of objects to space and of a behavior as a continuous function from time to layouts. To do so, however, would require reasoning about sets of objects, function spaces, and continuity. We have therefore short-circuited the process by postulating the results of direct interest.

**Axiom 2.18:** Two behaviors $B1$ and $B2$ which have equal layouts at some time $TS$ may be spliced together across time into a new behavior that agrees with $B1$ up to $TS$ and with $B2$ after $TS$.

$$\begin{array}{l} [\ \text{layout}(B1,TS) = \text{layout}(B2,TS)\ ] \Rightarrow \\ \exists_{B3}\ [\ \forall_{T \le TS}\ \text{layout}(B3,T) = \text{layout}(B1,T) \wedge \\ \qquad \forall_{T \ge TS}\ \text{layout}(B3,T) = \text{layout}(B2,T)\ ] \end{array}$$

**Axiom 2.19:** Any two behaviors $B1$ and $B2$ can be spliced together across objects to form a new behavior $B3$ which agrees with $B1$ on the objects of $B1$ and agrees with $B2$ on the objects that are in $B2$ but not in $B1$. Note that $B3$ may not be physically possible.

$$\begin{array}{l} \exists_{B3}\ \forall_{O,T} \\ \quad [\ \text{object\_of}(O,B1) \Rightarrow \\ \qquad [\ \text{object\_of}(O,B3) \wedge \text{place}(O,\text{layout}(B3,T)) = \text{place}(O,\text{layout}(B1,T))\ ]\ ] \wedge \\ \quad [\ [\ \text{object\_of}(O,B2) \wedge \neg\text{object\_of}(O,B1)\ ] \Rightarrow \\ \qquad [\ \text{object\_of}(O,B3) \wedge \text{place}(O,\text{layout}(B3,T)) = \text{place}(O,\text{layout}(B1,T))\ ]\ ] \end{array}$$

**Axiom 2.20:** Given a layout $L$ and an object $O$ in $L$, there exists a layout $LO$ which has $O$ in the same place as in $L$, and which has no other objects.

$$\text{object\_of}(O,L) \Rightarrow \exists_{LO}\ [\text{object\_of}(O1,LO) \Leftrightarrow O1 = O\ ] \wedge \text{place}(O,LO) = \text{place}(O,L)$$

**Definition 2.21:** The predicate "motionless($O,B,I$)" holds if the place occupied by $O$ in $B$ does not change during $I$. (By this definition, we are not counting rotation in place as motion.)

$$\begin{array}{l} \text{motionless}(O,B,I) \equiv \\ \forall_{T1,T2}\ [\ [\ \text{time\_in}(T1,I) \wedge \text{time\_in}(T2,I)\ ] \Rightarrow \\ \qquad \text{place}(O,\text{scene}(B,T1)) = \text{place}(O,\text{scene}(B,T2))\ ] \end{array}$$

**Definition 2.22:** The predicate "continual_motion($O,B,I$)" holds if $O$ is never motionless in $B$ throughout $I$.

$$\begin{array}{l} \text{continual\_motion}(O,B,I) \equiv \\ [\ \forall_{IS}\ \text{sub\_interval}(IS,I) \Rightarrow \neg\ \text{motionless}(O,B,IS)\ ] \end{array}$$

**Axiom 2.23:** Given any layout $L$, there is a behavior including $L$ in which everything is motionless.

$$\exists_B\ L = \text{layout}(B,TF) \wedge \forall_{I,O}\ \text{object\_of}(O,B) \Rightarrow \text{motionless}(O,B,I)$$

**Axiom 2.24:** If an object $O$ is strictly inside a region $XE$ in a layout $L$, then there exists a behavior $B$ which includes $L$, in which $O$ is in continual motion but stays inside $XE$.

$$\begin{array}{l} [\ \text{object\_of}(O,L) \wedge \text{strictly\_inside}(\text{place}(O,L),XE)\ ] \Rightarrow \\ \exists_B\ [\ L = \text{scene}(B,TF) \wedge \forall_T\ \text{strictly\_inside}(\text{place}(O,\text{scene}(B,T)),XE) \wedge \\ \qquad \forall_I\ \text{continual\_motion}(X,B,I)\ ] \end{array}$$

### 8.3. Physics

**Definition 3.1:** The predicate "object_of($O,L$)" holds if object $O$ is within layout $L$. The predicate "object_of($O,B$)" holds if object $O$ is within behavior $B$.

**Axiom 3.2:** A behavior has the same objects as each of its layouts.

$$\text{object\_of}(O,B) \Leftarrow \text{object\_of}(O,\text{scene}(B,T))$$

**Definition 3.3:** The function "place($O$,$L$)" maps an object $O$ and a layout $L$ into the point set occupied by $O$ during $L$.

**Definition 3.4:** The predicate "same_vprops($O1$,$O2$)" holds if $O1$ and $O2$ have the same visual properties.

**Axiom 3.5:** The relation "same_vprops($O1$,$O2$)" is an equivalence relation.

$$\begin{array}{l} \text{same\_vprops}(O,O) \wedge \\ \text{same\_vprops}(O1,O2) \Rightarrow \text{same\_vprops}(O2,O1) \\ [\ \text{same\_vprops}(O1,O2) \wedge \text{same\_vprops}(O2,O3)\ ] \Rightarrow \text{same\_vprops}(O1,O3) \end{array}$$

**Definition 3.6:** The predicate "free_space($X$,$L$)" holds if point set $X$ does not intersect the place of any bodies in layout $L$.

$$\text{free\_space}(X,L) \Leftrightarrow [\ \forall_O\ \text{object\_of}(O,L) \Rightarrow \neg\text{intersect}(X,\text{place}(O,L))\ ]$$

**Definition 3.7:** The predicate "visible($P$,$A$,$L$)" holds if the point $P$ is visible to agent $A$ in layout $L$. That is, there is no object between $A$ and $P$. (Note that each point in $A$ itself is visible to $A$ by definition.)

$$\begin{array}{l} \text{visible}(P,A,L) \Leftrightarrow \\ \forall_O\ [\ [\text{object\_of}(O,L) \wedge O \neq A\ ] \Rightarrow \\ \quad \exists_{PA}\ \text{point\_in}(PA,\text{place}(A,L)) \wedge \neg\text{blocked}(P,\text{place}(O,L),PA)\ ] \end{array}$$

**Axiom 3.8:** If a point $P$ is invisible to agent $A$ in $L$, then, for any point $PA$ in $A$, there is some visible point on an object that blocks $P$ from $PA$. (This is non-trivial; it relies on an assumption that there are only finitely many objects in any finite area. Imagine looking down at an infinite stack of pages, each half as thin as the one underneath. Then none of the pages would be visible.)

$$\begin{array}{l} [\ \neg\text{visible}(P,A,L) \wedge \text{point\_in}(PA,\text{place}(A,L))\ ] \Rightarrow \\ \exists_O\ [\ \text{blocked}(P,\text{place}(O,L),PA)) \wedge \\ \quad \exists_{PO}\ [\text{point\_in}(PO,\text{place}(O,L)) \wedge \text{visible}(PO,A,L) \wedge \\ \qquad \text{point\_in}(PO,\text{line\_seg}(PA,P))\ ]\ ] \end{array}$$

**Definition 3.9:** The predicate "wholly_visible($X$,$A$,$L$)" holds if each point of $X$ is visible to $A$ in $L$.

$$\text{wholly\_visible}(X,A,L) \Leftrightarrow [\ \forall_P\ \text{point\_in}(P,X) \Rightarrow \text{visible}(P,A,L)\ ]$$

**Definition 3.10:** The predicate "wholly_invisible($X$,$A$,$L$)" holds if no point of $X$ is visible to $A$ in $L$.

$$\text{wholly\_invisible}(X,A,L) \Leftrightarrow [\ \forall_P\ \text{point\_in}(P,X) \Rightarrow \neg\text{visible}(P,A,L)\ ]$$

**Definition 3.11:** The predicate "phys_poss($L$)" holds if $L$ is a physically possible layout.

**Axiom 3.12:** A layout is physically possible if no two objects overlap.

$$\begin{array}{l} \text{phys\_poss}(L) \Leftrightarrow \\ [\ \forall_{O1,O2}\ [\text{object\_of}(O1,L) \wedge \text{object\_of}(O2,L)\ ] \Rightarrow \\ \qquad \neg\text{overlap}(\text{place}(O1,L),\text{place}(O2,L))\ ] \end{array}$$

**Axiom 3.13:** A behavior is physically possible if each layout is physically possible.

$$\text{phys\_poss}(B) \Leftrightarrow \forall_T\ \text{phys\_poss}(\text{scene}(B,T))$$

**Axiom 3.14:** Any chronicle is physically possible.

phys_poss(behavior($C$))

### 8.4. Perception and Knowledge

**Definition 4.1:** The predicate "v_compatible($A$ ,$L1$,$L2$)" means that layout $L2$ is visually compatible with layout $L1$ relative to agent $A$. That is, as far as $A$ can see in $L1$, the world might be in state $L2$.

**Axiom 4.2:** Layout $L2$ is v_compatible with $L1$ with respect to agent $A$ iff the following two conditions hold: (i) $L1$ and $L2$ are both physically possible. (ii) Let $X$ be a connected set of points, such that every point in $X$ is visible to $A$ in $L1$. Then each point of $X$ is visible to $A$ in $L2$. Moreover, if $X$ lies entirely in $A$ in $L1$, then $X$ lies entirely in $A$ in $L2$; if $X$ lies entirely in some object $O1$ in $L1$ then, in $L2$, $X$ lies entirely within some object $O2$ with the same visible properties as $O1$; if $X$ lies entirely in free space in $L1$ then $X$ lies entirely in free space in $L2$.

$$
\begin{array}{l}
\text{v\_compatible}(A,L1,L2) \Leftrightarrow \\
\quad [\ \text{object\_of}(A,L1) \wedge \text{object\_of}(A,L2) \wedge \text{phys\_poss}(L1) \wedge \text{phys\_poss}(L2) \wedge \\
\quad\quad \forall_{X,O}\ [\ \text{connected}(X) \text{ and wholly\_visible}(X,A,L1)\ ] \Rightarrow \\
\quad\quad\quad [\ \text{wholly\_visible}(X,A,L2) \wedge \\
\quad\quad\quad\quad [\ \text{sub\_place}(X,\text{place}(A,L1)) \Rightarrow \text{sub\_place}(X,\text{place}(A,L2))\ ] \wedge \\
\quad\quad\quad\quad [\ [\ \text{object\_of}(O,L1) \wedge \text{sub\_place}(X,\text{place}(O,L1))\ ] \Rightarrow \\
\quad\quad\quad\quad\quad [\ \exists_{O2}\ \text{object\_of}(O2,L2) \wedge \text{same\_vprops}(O,O2) \wedge \text{sub\_place}(X,\text{place}(O2,L2))\ ]\ ] \wedge \\
\quad\quad\quad\quad [\ [\ \text{object\_of}(O,L2) \wedge \text{sub\_place}(X,\text{place}(O,L2))\ ] \Rightarrow \\
\quad\quad\quad\quad\quad [\ \exists_{O1}\ \text{object\_of}(O1,L1) \wedge \text{same\_vprops}(O,O1) \wedge \text{sub\_place}(X,\text{place}(O1,L1))\ ]\ ] \wedge \\
\quad\quad\quad\quad [\ \text{free\_space}(X,L1) \Rightarrow \text{free\_space}(X,L2)\ ]\ ]\ ]
\end{array}
$$

**Lemma 4.3:** v_compatibility is an equivalence relation on layouts.

$$
\begin{array}{l}
\text{v\_compatible}(A,L,L) \\
\text{v\_compatible}(A,L1,L2) \Rightarrow \text{v\_compatible}(A,L2,L1) \\
[\ \text{v\_compatible}(A,L1,L2) \wedge \text{v\_compatible}(A,L2,L3)\ ] \Rightarrow \text{v\_compatible}(A,L1,L3)
\end{array}
$$

**Proof:** Reflexivity and transitivity are immediate from axiom 4.2, since the right-hand side of the definition has throughout the form "Property of L1 implies same property of L2." This rests on the fact that same_vprops is an equivalence relation (axiom 3.5). The axiom is also nearly symmetric in $L1$ and $L2$: the only missing point is to show that, if $L1$ and $L2$ satisfy the condition of the axiom, and $X$ is wholly visible in $L2$ then it must also be wholly visible in $L1$. This can be proved by contradiction as follows: Let $L1$ and $L2$ be two layouts satisfying the right-hand side of axiom 4.2, and let $X$ be a connected set of points visible to $A$ in $L2$. Suppose that $X$ were not wholly visible in $L1$, so that (by definition 3.9) some point $PX$ in $X$ were invisible in $L1$. Let $PA$ be a point in $A$. From axiom 3.8, we know that there is in $L1$ some object $O1$ with a visible part $XO$ that blocks $PX$ from $PA$. By the right-hand side of axiom 4.2, in $L2$, $XO$ must lie in some object $O2$. But then $O$ must block $PX$ from $PA$ in $L2$. Since $PA$ was chosen arbitrarily, $PX$ must be blocked from every point in $A$, which means (by definition 3.9) that $X$ is not wholly visible in $L2$. This establishes the contradiction.

**Lemma 4.4:** Let $L1$ and $L2$ be physically possible layouts with the following properties: (i) The same objects are partly visible in the two layouts. (ii) Any object which is partly visible in the two layouts is in the same position in both. Then the two layouts are visually compatible.

$$
\begin{array}{l}
[\ \text{object\_of}(A,L1) \wedge \text{object\_of}(A,L2) \wedge \text{phys\_poss}(L1) \wedge \text{phys\_poss}(L2) \wedge \\
\quad [\ [\ \text{object\_of}(O,L1) \wedge \neg\text{wholly\_invisible}(O,A,L1)\ ] \Rightarrow \\
\qquad [\ \text{object\_of}(O,L2) \wedge \text{place}(O,L1) = \text{place}(O,L2)\ ]\ ] \wedge \\
\quad [\ [\ \text{object\_of}(O,L2) \wedge \neg\text{wholly\_invisible}(O,A,L2)\ ] \Rightarrow \\
\qquad [\ \text{object\_of}(O,L1) \wedge \text{place}(O,L1) = \text{place}(O,L2)\ ]\ ] \\
] \Rightarrow \text{v\_compatible}(A,L1,L2)
\end{array}
$$

**Proof:** Assume that the antecedent of the lemma holds for some layouts $L1$ and $L2$. Note that, since $A$ is always wholly visible to itself, it must occupy the same place in $L1$ as in $L2$. Let $X$ be a set of points which is wholly visible in $L1$. We will show (A) that $X$ is wholly visible in $L2$ and (B) that $X$ is inside "the same kind of space" in $L2$ as in $L1$. Using axiom 4.2, it will follow that $L1$ and $L2$ are visually compatible.

A. The proof by contradiction is almost the same as that of lemma 4.3: Suppose that $X$ were not wholly visible in $L2$, so that (by definition 3.9) some point $PX$ in $X$ were invisible in $L2$. Let $PA$ be a point in $A$. From axiom 3.8, we know that there is in $L2$ some visible object $O$ which blocks $PX$ from $PA$. By the antecedent of the lemma, $O$ must occupy the same place in $L1$. But then $O$ must block $PX$ from $PA$ in $L1$. Since $PA$ was chosen arbitrarily, $PX$ must be blocked from every point in $A$, which means (by definition 3.9) that $X$ is not wholly visible in $L1$. This contradicts the hypothesis of the proof.

B. (i) Let $X$ be a subset of $A$ in $L1$. Since $A$ occupies the same space in $L1$ as in $L2$, $X$ must be a subset of $A$ in $L2$.

(ii) Let $X$ be a non-null subset of $O$ in $L1$. Since $X$ is wholly visible, $O$ is not wholly invisible (Definitions 1.3, 3.9, 3.10.) Hence, $O$ occupies the same space in $L2$, and $X$ must be a subset of $O$ in $L2$. Note that, by axiom 3.5, $O$ has the same visual properties as itself.

(iii) Let $X$ be in free space in $L1$. Suppose that $X$ is not in free space in $L2$; that is (definition 3.6), $X$ intersects some object $O$ in $L2$. Since $X$ is visible in $L2$, $O$ must be partially visible, so, by the antecedent of the lemma, $O$ must be in the same place in $L1$ as in $L2$. But then $X$ is not in free space in $L1$, contrary to hypothesis.

Thus all the conditions of axiom 4.2 are satisfied, completing the proof.

**Definition 4.5:** The predicate "bv_compatible($A$,$B1$,$B2$,$T$)" means that behavior $B2$ is visually compatible with behavior $B1$ up to time $T$. That is, $B2$ is consistent with everything that $A$ can see in $B1$ up to (and including) time $T$.

**Axiom 4.6:** Two behaviors are visually compatible up to time $T$ if corresponding layouts up to time $T$ are compatible.

$$
\begin{array}{l}
\text{bv\_compatible}(A,B1,B2,TS) \Leftarrow \\
\forall_{T \le TS}\ \text{v\_compatible}(A,\text{situation}(B1,T),\text{situation}(B2,T))
\end{array}
$$

**Lemma 4.7:** bv_compatibility is an equivalence relation over behaviors.

$$
\begin{array}{l}
\text{bv\_compatible}(A,B,B,T). \\
\text{bv\_compatible}(A,B1,B2,T) \Rightarrow \text{bv\_compatible}(A,B2,B1,T). \\
[\ \text{bv\_compatible}(A,B1,B2,T) \wedge \text{bv\_compatible}(A,B2,B3,T)\ ] \Rightarrow \\
\qquad \text{bv\_compatible}(A,B1,B3,T).
\end{array}
$$

**Proof:** Immediate from lemma 4.3 and axiom 4.6.

**Definition 4.8:** The predicate "k($A$,$S1$,$S2$)" holds if $S2$ is accessible in $S1$ via $A$'s knowledge. That is, $S2$ is consistent with everything that $A$ knows in $S1$.

**Axiom 4.9:** Veridicality: All that is known is true. Formally, the knowledge accessibility relation is reflexive.

$$\text{k}(A,S,S)$$

**Axiom 4.10:** Positive introspection: If $A$ knows $\phi$, then he knows that he knows $\phi$. Formally, the knowledge accessibility relation is transitive.

$$[\ k(A, S1, S2) \wedge k(A, S2, S3)\ ] \Rightarrow k(A, S1, S3)$$

**Axiom 4.11:** Memory: An agent does not forget what he knows. Formally, let situation $S1B$ be accessible from $S0B$, and let $S0A$ precede $S0B$ in the same chronicle. Then there is a situation $S1A$ in the chronicle of $S1B$ which is accessible from $S0A$.

$$[\ k(A, S0B, S1B) \wedge \text{precedes}(S0A, S0B)\ ] \Rightarrow$$
$$\exists_{S1A}\ [\ k(A, S0A, S1A) \wedge \text{precedes}(S1A, S1B)\ ]$$

**Axiom 4.12:** Internal clock: An agent always knows the time. Formally, if $S1$ is knowledge accessible from $S0$, then the two situations occur at the same time.

$$k(A, S0, S1) \Rightarrow \text{time}(S0) = \text{time}(S1)$$

**Axiom 4.13:** An agent knows what he perceives. Formally if $S1$ is knowledge accessible from $S0$, then the layout of $S1$ is visually compatible with the layout of $S0$.

$$k(A, S0, S1) \Rightarrow \text{v\_compatible}(A, \text{layout}(S0), \text{layout}(S1))$$

**Axiom 4.14:** Perception is the only source of knowledge of the course of events. Formally, let $C0$ be the real chronicle. Let $B1$ be a behavior that is visually compatible with the behavior of $C0$ up to time $T$ relative to agent $A$; thus, as far as $A$ could have seen up to time $T$, $B1$ could be the real behavior. Then it is consistent with $A$'s knowledge that $B1$ actually was the real behavior; that is, there is a chronicle $C1$ whose situation in $T$ is knowledge accessible from the situation of $C0$ in $T$.

$$\text{bv\_compatible}(A, \text{behavior}(C0), B1, T) \Rightarrow$$
$$\exists_{C1}\ [\ k(A, \text{situation}(C0,T), \text{situation}(C1,T)) \wedge B1 = \text{behavior}(C1)\ ]$$

**Theorem 4.15:** Facts perceived over time are known. Formally, if $S1$ is knowledge accessible from $S0$, then the behavior of the chronicle of $S0$ up $S0$ is visually compatible with the behavior of the chronicle of $S1$ up to $S1$.

$$k(A, S0, S1) \Rightarrow$$
$$\text{bv\_compatible}(A, \text{behavior}(\text{chronicle}(S0)), \text{behavior}(\text{chronicle}(S1)), \text{time}(S0))$$

**Proof:** Let $S1$ be accessible from $S0$. Let $C0$ be the chronicle of $S0$, let $C1$ be the chronicle of $S1$, and let $TS$ be the time of both $S0$ and $S1$. (By axiom 4.12, they must have equal times.) By axiom 4.11, of memory, for any situation $S0A$ preceding $S0$ there must be a situation $S1A$ preceding $S1$ which is accessible from $S0A$. By axiom 4.12, that the agent knows the time, these must have the same time. Thus, using definition 2.13, we have the following statement: If $T \leq TS$ then situation($C1,T$) is accessible from situation($C2,T$). By axiom 4.13, that perceptions are known, this means that the layouts layout(situation($C0,T$)) and layout(situation($C1,T$)) are visually compatible. By axiom 2.17 these layouts are respectively equal to scene(behavior($C0$),$T$), and scene(behavior($C1$),$T$). Thus, the two behaviors behavior($C0$) and behavior($C1$) have visually compatible layouts at each time $T \leq TS$, so they are bv_compatible up to time $TS$.

**Theorem 4.16:** Negative introspection: If an agent does not know a fact about the course of events, then he knows that he doesn't know it. Formally, if a behavior is compatible with an agent's perceptions, then, from any knowledge accessible situation, there is a knowledge accessible chronicle exhibiting that behavior.

$$[\ \text{bv\_compatible}(A, \text{behavior}(C0), B1, T) \wedge k(A, \text{situation}(C0,T), SM)\ ] \Rightarrow$$
$$\exists_{C1}\ [\ k(A, SM, \text{situation}(C1,T)) \wedge \text{bv\_compatible}(A, \text{behavior}(C1), B1, T)\ ]$$

**Proof:** Let $A$, $B1$, $C0$, $SM$, and $T$ satisfy the antecedent of the theorem. Let $CM$ be the chronicle of $SM$. By theorem 4.15, the behavior of $CM$ is visually compatible with the behavior of $C0$. Since, by lemma 4.7, bv_compatibility is an equivalence relation, it follows that $B1$ is compatible with the behavior of $CM$. Therefore, by axiom 4.14, $A$ cannot know in $SM$ that $B1$ has not occurred; that is, there exists a $C1$ with a situation accessible from $SM$ whose behavior is $B1$.

## 8.5. Example 1: Inferring Ignorance

The basic constants of our example are the following:

Constants:

| | | |
|---|---|---|
| aclaire | — | Claire |
| c0 | — | Actually occurring chronicle |
| i0 | — | Time interval in question |
| omystery | — | The mystery object. |
| owall | — | The wall |
| xclaire | — | Place occupied by Claire |
| xenvelope | — | The spatial envelope containing the object. |
| xwall | — | Place occupied by wall |

We define a few additional constants as convenient abbreviations:

**Definition 5.1:**

| | | | | |
|---|---|---|---|---|
| b0 | ≡ | behavior(c0) | — | The real behavior. |
| ta | ≡ | start(i0) | — | Starting time |
| tz | ≡ | end(i0) | — | Starting time |
| la | ≡ | scene(b0,ta) | — | Starting layout |
| s0z | ≡ | situation(c0, tz) | — | The ending situation |

**Hypothesis 5.2:** Claire, the wall, and the mystery object are distinct objects in the chronicle.

object_of(aclaire,b0) $\wedge$ object_of(owall,b0) $\wedge$ object_of(omystery,b0) $\wedge$
aclaire $\neq$ owall $\wedge$ owall $\neq$ omystery $\wedge$ aclaire $\neq$ omystery

**Hypothesis 5.3:** Claire occupies the fixed place xclaire throughout the interval i0.

time_in($T$,i0) $\Rightarrow$ place(aclaire, scene(b0,$T$)) = xclaire

**Hypothesis 5.4:** The wall occupies the fixed place xwall throughout the interval i0.

time_in($T$,i0) $\Rightarrow$ place(owall, scene(b0, $T$)) = xwall

**Hypothesis 5.5:**

The mystery object remains strictly inside the envelope throughout the interval i0.

time_in($T$,i0) $\Rightarrow$ strict_inside(place(omystery,scene(b0,$T$)), xenvelope)

**Hypothesis 5.6:** No other object comes inside the envelope within the interval i0.

[ $O \neq$ omystery $\wedge$ time_in($T$,i0) ] $\Rightarrow$ $\neg$overlap(place($O$, scene(b0,$T$)), xenvelope)

**Hypothesis 5.7:** The envelope is blocked by the wall from Claire.

blocked(xenvelope, xwall, xclaire)

**To prove:** Claire does not know whether or not omystery was motionless during i0. We express this as follows: There is a situation accessible to Claire in situation s0z (the ending

situation) that follows on a chronicle in which omystery was motionless throughout i0. There is also an accessible situation that follows on a chronicle in which it was not motionless.

$$[\ \exists_{C1}\ k(\text{aclaire},\text{s0z},\text{situation}(C1,\text{tz})) \wedge \text{motionless}(\text{omystery},\text{behavior}(C1),\text{i0})\ ] \wedge$$
$$[\ \exists_{C2}\ k(\text{aclaire},\text{s0z},\text{situation}(C2,\text{tz})) \wedge \neg\text{motionless}(\text{omystery},\text{behavior}(C2),\text{i0})\ ]$$

**Lemma 5.9:** There is a behavior, which we henceforth call "b1", containing the same objects as b0, such that (i) all layouts of b1 are the same as those of b0, up to and including time ta; (ii) every object except omystery has the same behavior in b1 as in in b0; (iii) omystery is motionless after time ta.

$$\begin{aligned}
&\exists_{B1}\ \forall_O\ [\ \text{object_of}(O,B1) \Leftrightarrow \text{object_of}(O,\text{b0})\ ] \wedge \\
&\qquad \forall_{T \le ta}\ \text{layout}(B1,T) = \text{layout}(\text{b0},T) \wedge \\
&\qquad [\ [\text{object_of}(O,B1) \wedge O \ne \text{omystery}\ ] \Rightarrow \\
&\qquad\quad \forall_T\ \text{place_of}(O,\text{scene}(B1,T)) = \text{place_of}(O,\text{scene}(\text{b0},T))\ ] \wedge \\
&\qquad \forall_I\ [\text{ta} \le \text{start}(I) \Rightarrow \text{motionless}(\text{omystery},B1,I)\ ]
\end{aligned}$$

**Proof:** By axiom 2.20, there is a layout *LAP* whose only object is omystery, placed in the same place as in the layout "la". By axiom 2.23, there is a behavior *BX* containing *LAP* at time ta with omystery motionless throughout *BX*. By axiom 2.19, there is a behavior *BY* combining the behavior of omystery from *BX* and the behavior of the other objects from b0. From the construction of *BY*, and from axiom 2.9, it follows that the layout of *BY* at time ta is equal to the layout la. Therefore, by axiom 2.18, we can define b1 to be the behavior which agrees with b0 up to time ta, and with *BY* after ta. The above properties follow directly from the construction.

**Lemma 5.10:** There is a behavior, which we henceforth call "b2", containing the same objects as b0, such that (i) all layouts of b2 are the same as those of b0, up to and including time ta; (ii) every object except omystery has the same behavior in b2 as in in b0; (iii) after time ta, omystery stays inside xenvelope in a state of continual motion.

$$\begin{aligned}
&\exists_{B1}\ \forall_O\ [\ \text{object_of}(O,B1) \Leftrightarrow \text{object_of}(O,\text{b0})\ ] \wedge \\
&\qquad \forall_{T \le ta}\ \text{layout}(B1,T) = \text{layout}(\text{b0},T) \wedge \\
&\qquad [\ [\text{object_of}(O,B1) \wedge O \ne \text{omystery}\ ] \Rightarrow \\
&\qquad\quad \forall_T\ \text{place_of}(O,\text{scene}(B1,T)) = \text{place_of}(O,\text{scene}(\text{b0},T))\ ] \wedge \\
&\qquad \forall_{T \ge ta}\ \text{strictly_inside}(\text{place}(\text{omystery},\text{scene}(B2,T)),\text{xenvelope}) \wedge \\
&\qquad \forall_I\ [\text{ta} \le \text{start}(I) \Rightarrow \text{continual_motion}(\text{omystery},B1,I)\ ]
\end{aligned}$$

**Proof:** Exactly the same as that of lemma 5.9, except that, instead of using axiom 2.23 to construct a behavior in which omystery stands still, we use axiom 2.24 together with hypothesis 5.5 to construct a behavior in which omystery is in continuous motion within xenvelope.

**Lemma 5.11:** In behavior b1 at all times after ta, omystery is inside xenvelope.

$$\forall_{T \ge ta}\ \text{strictly\_inside}(\text{place}(\text{omystery},\text{scene}(\text{b1},T)),\ \text{xenvelope})$$

**Proof:** By hypothesis 5.5, omystery is inside xenvelope in layout la. By lemma 5.9, the layout of b1 at ta is equal to la. By lemma 5.9, o1 is motionless in b1 in all intervals starting after or at time ta. By definition 2.21 and axiom 2.6, this means that o1 always occupies the same place, inside xenvelope, at all times after ta.

**Lemma 5.12:** Behaviors b1 and b2 is both physically possible.

$$\text{phys\_poss}(\text{b1}) \wedge \text{phys\_poss}(\text{b2})$$

**Proof:** The argument is almost identical for both b1 and b2; we will use b1 for illustration. By axiom 3.13, it suffices to show that each layout scene(b1,$T$) is physically possible. By axiom 2.2, any time $T$ is either before or after ta. By lemma 5.9, for $T \le$ ta, scene(b1,$T$)

= scene(b0,$T$), and by axioms 3.13 and 3.14 the layouts of b0 are physically possible. To show that scene(b1,$T$) is physically possible, it suffices, by axiom 3.12, to show that the places of no two objects overlap in the layout. If neither of the two objects is equal to omystery, then, by lemma 5.9, their places in b1 are the same as in b0. Since, by axioms 3.12, 3.13, and 3.14, they do not overlap in b0, they cannot overlap in b1.

The remaining case to consider is to rule out the possibility that omystery overlaps some other object, say o1, after time ta. By lemma 5.9 for b1, and by lemma 5.10 for b2, after time ta, omystery is strictly inside the region xenvelope; hence, by definition 1.11, the place of omystery is a subset of xenvelope. By hypothesis 5.6, o1 never occupies a place overlapping xenvelope. Hence, by lemma 1.9, the place of o1 never overlaps the place of omystery.

**Lemma 5.13:** In each of the behaviors b0, b1, and b2, at all times during the interval i0, the object omystery is invisible to claire.

$$[\,[\,B = \text{b1} \vee B = \text{b2} \vee B = \text{b3}\,] \wedge L = \text{scene}(B,T)\,] \Rightarrow$$
$$\text{wholly\_invisible(omystery, aclaire, } L)$$

**Proof:** From hypotheses 5.3, 5.4, and 5.5 for b0, lemma 5.9 for b1, and lemma 5.10 for b2, the mystery object remains in the point set xenvelope throughout i0: Claire remains in xclaire; and the wall remains filling xwall. By hypothesis 5.7, xwall blocks xenvelope from xclaire. By lemma 1.17, therefore, the place of the mystery object is always blocked by xwall from xclaire. Hence, using definitions 3.7 and 3.10 and hypothesis 5.1, the mystery object is always invisible to Claire.

**Lemma 5.14:** The behaviors b1 and b2 are each visually compatible with Claire's perceptions in b0 up to the end of i0.

$$\text{bv\_compatible(aclaire,behavior}(C)\text{,b1,tz)} \wedge \text{bv\_compatible(aclaire,behavior}(C)\text{,b2,tz)}$$

**Proof:** At all times up to the start of i0, b1 and b2 are identical to b0, by lemmas 5.9 and 5.10. During i0, by lemmas 5.9 and 5.10, the only object which may be in a different position in either b1 or b2 from its position in b0 is omystery, and by lemma 5.13, it is always wholly invisible to Claire throughout b0, b1, and b2. Moreover, by lemma 5.12, b1 and b2 are physically possible. Hence, from lemma 4.4, we may conclude that b1 and b2 are visually compatible with b0 up to time tz.

**Lemma 5.15:** The behavior b1 is visually compatible with Claire's perceptions up to the end of i0, and omystery is motionless in b1 during i0.

$$\text{bv\_compatible(aclaire,behavior}(C)\text{,b1,end(i0))} \wedge \text{motionless(omystery,b1,i0)}\,]$$

**Proof:** Immediate from lemmas 5.9 and 5.14

**Lemma 5.16:** The behavior b2 is visually compatible with Claire's perceptions up to the end of i0, and omystery is not motionless in b2 during i0.

$$\text{bv\_compatible(aclaire,behavior}(C)\text{,b1,tz)} \wedge \neg\text{motionless(omystery,b1,i0)}\,]$$

**Proof:** Immediate from lemmas 5.10 and 5.14

**Theorem 5.17:** Claire does not know, at the end of i0, whether or not omystery has been motionless during i0.

$$[\,\exists_{C1}\ \text{k(aclaire,s0z, situation}(C1\text{,tz))} \wedge \text{motionless(omystery,behavior}(C1)\text{,i0)}\,] \wedge$$
$$[\,\exists_{C2}\ \text{k(aclaire,s0z,situation}(C2\text{,tz))} \wedge \neg\text{motionless(omystery,behavior}(C2)\text{,i0)}\,]$$

**Proof:** Immediate from Axiom 4.14 and Lemmas 5.15 and 5.17.

## 9. References

[Appelt. 82] D. Appelt, "Planning Natural-Language Utterances to Satisfy Multiple Goals," SRI Technical Note 259, 1982.

[Allen and Perrault, 80] J. Allen and C. Perrault, "Analyzing intention in utterances," *Artificial Intelligence*, vol. 15, 1980, pp. 143-178.

[Barwise and Perry, 82] J. Barwise and J. Perry, *Situations and Attitudes*, MIT Press, 1982.

[Davis, 84] E. Davis, "An Ontology of Physical Action," NYU Tech. Rep. 123, 1984.

[Davis, 86a] E. Davis, *Representing and Acquiring Geographic Knowledge*, Pitman Publishing, 1986.

[Davis, 86b] E. Davis, "A Logical Framework for Solid Object Physics," NYU Tech. Rep. 245, 1986

[Hintikka, 69] J. Hintikka. "On the Logic of Perceptions," in *Models for Modalities* D. Reidel Publishing, Dordrechts, Holland, 1969.

[Hintikka, 71] J. Hintikka, "Semantics for Propositional Attitudes," in L. Linsky, ed. *Reference and Modality*, Oxford University Press, 1971.

[Konolige, 86] K. Konolige, *A Deduction Model of Belief*, Pitman Publishing, 1986.

[Levesque, 84] H. Levesque, "A Logic of Explicit and Implicit Belief," *Proc. AAAI*, 1984, pp. 198-202.

[McDermott, 82] D. McDermott, "A Temporal Logic for Reasoning about Processes and Plans," *Cognitive Science*, vol. 6, 1982, pp. 101-155.

[Moore, 80] R. Moore, *Reasoning about Knowledge and Action*, SRI Technical Note 191, 1980.

[Morgenstern, 87] L. Morgenstern, "Foundations of a Logic of Knowledge, Action, and Communication," NYU Ph.D. Thesis, 1987.

[Reiter and Mackworth, 87] R. Reiter and A. Mackworth, "The Logic of Depiction," Tech Rep. 87-24, U. British Columbia Computer Science Dept., 1987.

[Schank, 75] R. Schank, *Conceptual Information Processing*, North Holland, Amsterdam, 1975.

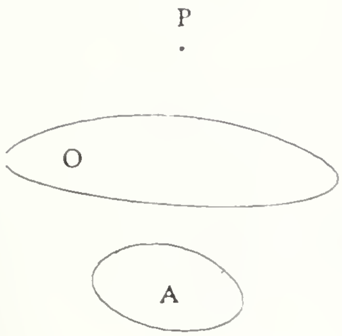



Figure 1:
O occludes P from A

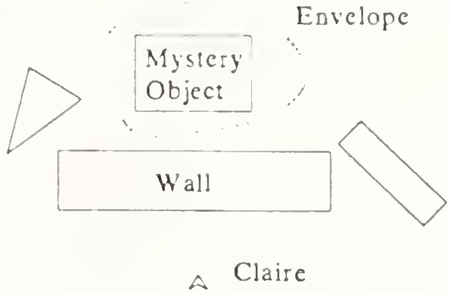



Figure 2:
Example I:
Claire cannot know
whether the object is moving

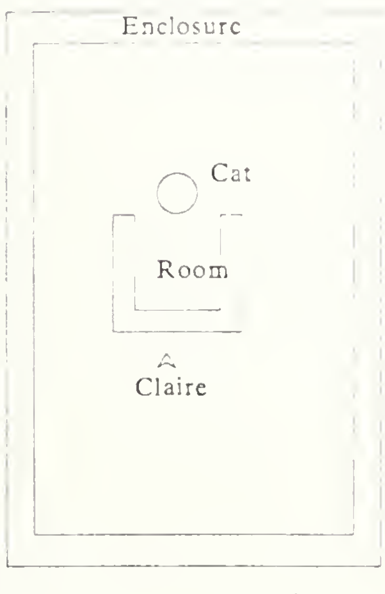



Figure 3:
Example III:
If Claire goes to the room
she will see whether
the cat is there

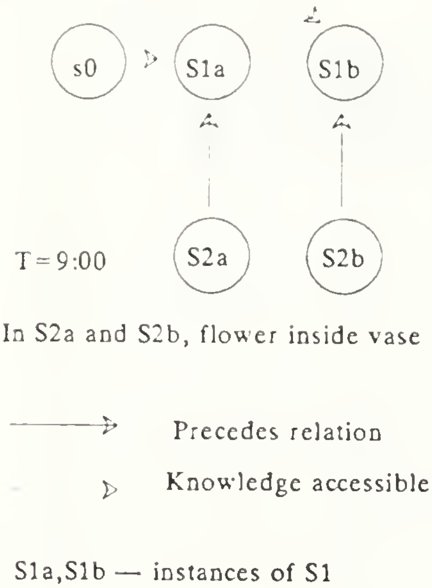



Figure 4:
Mark knows the flower was
in the vase at 9:00

If these relations hold

$$\begin{tikzcd} S0B \arrow[r] & S1B \\ S0A \arrow[u] & \end{tikzcd}$$

then the diagram can be completed

$$\begin{tikzcd} S0B \arrow[r] & S1B \\ S0A \arrow[u] \arrow[r] & S1A \arrow[u] \end{tikzcd}$$

Figure 5:
Axiom of Memory

NYU COMPSCI TR-328 c.1
Davis, Ernest
Inferring ignorance from
the locality of visual
perception.